What Did They Do in IEA 34/43? Or How to Diagnose and Repair Bugs in 500,000 Lines of Code

Preprint

R. Judkoff
National Renewable Energy Laboratory

J. Neymark
J. Neymark & Associates

To be presented at the Building Simulation 2009 Conference
Glasgow, Scotland
27–30 July 2009
NOTICE

The submitted manuscript has been offered by an employee of the Alliance for Sustainable Energy, LLC (ASE), a contractor of the US Government under Contract No. DE-AC36-08-GO28308. Accordingly, the US Government and ASE retain a nonexclusive royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for US Government purposes.

This report was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or any agency thereof.

Available electronically at http://www.osti.gov/bridge

Available for a processing fee to U.S. Department of Energy and its contractors, in paper, from:
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
phone: 865.576.8401
fax: 865.576.5728
email: reports@adonis.osti.gov

Available for sale to the public, in paper, from:
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
phone: 800.553.6847
fax: 703.605.6900
e-mail: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/ordering.htm

Printed on paper containing at least 50% wastepaper, including 20% postconsumer waste
WHAT DID THEY DO IN IEA 34/43?

OR HOW TO DIAGNOSE AND REPAIR BUGS IN 500,000 LINES OF CODE

Ron Judkoff¹ and Joel Neymark²

¹National Renewable Energy Laboratory, Golden, Colorado, USA
²Neymark and Associates, Golden, Colorado, USA

ABSTRACT

This paper presents an overview of the recently completed International Energy Agency Solar Heating and Cooling Task 34 and Energy Conservation in Buildings and Community Systems Annex 43 (IEA 34/43) on Testing and Validation of Building Energy Simulation Tools. The paper includes discussion of the technical and historical context, methodology, organization, accomplishments, and adoption by standards organizations and other regulatory entities.

INTRODUCTION

Modern whole-building energy simulation software may contain on the order of a half million lines of code. It is therefore helpful to develop testing and diagnostic methods that identify errors and indicate where in the code those errors are to facilitate corrections. The National Renewable Energy Laboratory (NREL) and the IEA have maintained a validation effort related to building energy simulation software since the first ECBCS project, Annex 1, which ran from 1977 to 1980 (ECBCS News, October 2007). Follow-on validation work was done in the IEA Solar Heating and Cooling Programme (SHC) Task 8, SHC Task 12/ECBCS Annex 21, and SHC Task 22. The most recent work was conducted under a combined effort of IEA SHC Task 34 and ECBCS Annex 43. The work began in June 2002 and was formally completed in December 2007. In all, 53 experts from 32 organizations and 13 countries contributed significantly to the Task in 6 project areas. These included A) Building to ground heat transfer B1) Multi-zone effects, not including airflow B2) Multi-zone airflow effects C) Shading, daylighting, and load interactions D) Hydronic heating and cooling systems, and E) Double-skin facades. The tests developed under these projects included empirical, comparative, and analytical solution-based methods, all of which are integral parts of the overall NREL/IEA/ASHRAE validation methodology, sometimes called BESTEST.

A new method based on stand-alone numerical solutions was developed that extends the analytical solution approach such that it can be applied to more realistic and less constrained test cases. The empirical validation work involved the construction and use of some unique test facilities designed expressly for the purpose of model validation. These included a double-façade test facility at Aalborg University in Denmark, Test Cells at EMPA in Switzerland, a test building at the Iowa Energy Center in the United States, and laboratory tests at the Technical University of Dresden in Germany. Participants in the Task either developed test suites or subjected building energy simulation software to the tests, or both, in an iterative process that facilitated the improvement of the test case specifications and the software. Each of the 6 full-length technical reports that resulted from the work describes the specifications for the tests in enough detail that an independent software producer or software user could perform the tests. In addition, these documents discuss the results from the models that were tested, the experiences of those who performed the tests, and improvements to software from the testing. Overall, 24 of the world's leading building simulation models were tested, resulting in more than 100 identified bugs and about 80 software corrections and/or improvements.

THEORY

There are three ways to evaluate a whole-building energy simulation program’s accuracy (Judkoff et al. 1983/2008; Judkoff and Neymark 2006):

- Empirical validation, which compares calculated results from a program, subroutine, algorithm, or software object to monitored data from a real building, test cell, or laboratory experiment.

¹This work was supported by the U.S. Department of Energy under Contract No. DE-AC36-08GO28308 with the National Renewable Energy Laboratory. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.
• **Analytical verification**, which compares outputs from a program, subroutine, algorithm, or software object to results from a known analytical solution or to results from a set of closely agreeing quasi-analytical solutions or verified numerical models

• **Comparative testing**, which compares a program to itself or to other programs. Table 1 compares these techniques (Judkoff et al. 1983/2008; Judkoff 1988; Judkoff and Neymark 2006). In this table, the term *model* is the representation of reality for a given physical behavior. For example, heat transfer may be simulated with one-, two-, or three-dimensional thermal conduction models. The term *solution process* encompasses the mathematics and computer coding to solve a given model. The solution process for a model can be perfect, even though the model remains inappropriate for a given physical situation, such as using a one-dimensional conduction model where two-dimensional conduction dominates. The term *truth standard* represents the standard of accuracy for predicting real behavior. An analytical solution is a “mathematical truth standard,” and tests the solution process for a model, but not the appropriateness of the model. An approximate truth standard from an experiment tests both the solution process and appropriateness of the model within experimental uncertainty. The ultimate validation truth standard would be comparison of simulation results with a perfectly performed empirical experiment, with all simulation inputs perfectly defined.

### Table 1
**Validation Techniques**

<table>
<thead>
<tr>
<th>TECHNIQUE</th>
<th>ADVANTAGES</th>
<th>DISADVANTAGES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Empirical Validation</strong></td>
<td>• Approximate truth standard within experimental accuracy</td>
<td>• Experimental uncertainties: - Instrument calibration, and spatial/temporal discretization - Imperfect knowledge/specification of experimental object (building) being simulated</td>
</tr>
<tr>
<td>Test of model and solution process</td>
<td>• Any level of complexity</td>
<td>• High-quality, detailed measurements are expensive and time-consuming • Only a limited number of test conditions are practical</td>
</tr>
<tr>
<td><strong>Analytical Verification</strong></td>
<td>• No input uncertainty</td>
<td>• No test of model validity • Limited to highly constrained cases for which analytical or quasi-analytical solutions can be developed</td>
</tr>
<tr>
<td>Test of solution process</td>
<td>• Exact mathematical or secondary mathematical truth standard for given model</td>
<td></td>
</tr>
<tr>
<td><strong>Comparative Testing</strong></td>
<td>• No input uncertainty</td>
<td>No absolute truth standard (only statistically based acceptance ranges are possible)</td>
</tr>
<tr>
<td>Relative test of model and solution process</td>
<td>• Any level of complexity • Many diagnostic comparisons possible • Inexpensive and quick</td>
<td></td>
</tr>
</tbody>
</table>

*Source: Judkoff and Neymark 2006.*

Table 1 shows that each validation technique has different strengths and weaknesses. In practice it is very expensive to empirically validate objects of the scale and complexity of buildings. This limits high-quality empirical validation data to a very few cases and therefore limits the diagnostic power of the cases (we discuss high-quality validation data toward the end of this section). By combining empirical, analytical, and comparative validation techniques, we can compensate for the weaknesses in any individual technique.

A comparison between measured and calculated performance represents a small region in an immense \(N\)-dimensional parameter space. Investigators are constrained to exploring relatively few domains in this space, yet would like to be assured that the results are not coincidental (e.g., not a result of offsetting errors) and do represent the validity of the simulation elsewhere in the parameter space. Analytical and comparative techniques minimize the uncertainty of extrapolations around the limited number of sampled empirical domains. Table 2 classifies these extrapolations. Use of the term *vice versa* in Table 2 is intended to mean that the extrapolation can go both ways (e.g., from short-term to long-term data and from long-term to short-term data). This does not mean that such extrapolations are correct, but only that researchers and practitioners have either explicitly or implicitly made such inferences in the past.
Figure 1 shows one process to combine analytical, empirical, and comparative techniques. These three techniques may also be used together in other ways; for example, intermodel comparisons may be done before an empirical validation exercise, to better define the experiment and to help estimate experimental uncertainty by propagating all known error sources through one or more whole-building energy simulation programs (Hunn et al. 1982; Lomas et al. 1994).

Table 2
Types of Extrapolation

<table>
<thead>
<tr>
<th>OBTAINABLE DATA POINTS</th>
<th>EXTRAPOLATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>A few climates</td>
<td>Many climates</td>
</tr>
<tr>
<td>Short-term total energy use</td>
<td>Long-term total energy use, or vice versa</td>
</tr>
<tr>
<td>Short-term (hourly) temperatures and/or fluxes</td>
<td>Long-term total energy use, or vice versa</td>
</tr>
<tr>
<td>A few equipment performance points</td>
<td>Many equipment performance points</td>
</tr>
<tr>
<td>A few buildings representing a few sets of variable and parameter combinations</td>
<td>Many buildings representing many sets of variable and parameter combinations, or vice versa</td>
</tr>
<tr>
<td>Small-scale: simple test cells, buildings, and mechanical systems; laboratory experiments</td>
<td>Large-scale complex buildings with complex HVAC systems, or vice versa</td>
</tr>
</tbody>
</table>

Source: Neymark and Judkoff 2002

For the path shown in Figure 1, there are three steps:

1. Run the code against analytical verification test cases to check its mathematical solution. Discrepancies must be corrected before proceeding further.
2. Run the code against high-quality empirical validation data and correct any errors.
3. Check the agreement of several programs with different thermal solution and modeling approaches (that have passed through steps 1 and 2) in a variety of representative cases. This uses the comparative technique as an extrapolation tool. Deviations in the program predictions indicate areas for further investigation.

When programs successfully complete these three stages, they are considered validated for cases where acceptable agreement was achieved (i.e., for the range of building, climate, and mechanical system types represented by the test cases). Once several detailed simulation programs have satisfactorily completed the procedure, other programs and simplified design tools can be tested against them. A validated code does not necessarily represent truth. It does represent a set of algorithms that have been shown, through a repeatable procedure, to perform according to the current state of the art.

The NREL methodology for validating building energy simulation programs has been generally accepted by the International Energy Agency (Irving 1988), ASHRAE Standard 140 and Addendum p to ASHRAE Standard 90.1, and elsewhere.

We are frequently asked why high-quality empirical validation is so difficult? The simplest level of empirical validation compares a building’s long-term energy use to that calculated by a computer program, and in principle, any building with a meter could be used. Unfortunately, it is impossible to interpret the results of such an exercise because all possible sources of error are acting simultaneously. Even if good agreement between measured and calculated energy consumption is observed, possible offsetting errors prevent a definitive conclusion about a model’s accuracy. More informative levels of empirical validation involve controlling known sources of error to identify and quantify unknown error sources and to reveal causal relationships associated with error sources. These error sources may be separated into two groups.
External Error Types:

- Differences between actual building microclimate versus the weather input to the computer program
- Differences between actual schedules, control strategies, effects of occupant behavior, and other effects from the real building versus those assumed by the program user
- User error deriving building input files
- Differences between actual physical properties of the building (including HVAC systems) versus those input by the user.

Internal Error Types:

- Differences between actual thermal transfer mechanisms in the real building and its HVAC systems versus the simplified model of those processes in the simulation (all models, no matter how detailed, are simplifications of reality)
- Errors or inaccuracies in the mathematical solution of the models
- Coding errors.

Designing and conducting validation experiments that allow control or isolation of these various error sources is challenging and expensive, but not impossible. The best validation experiments are from test cells or buildings constructed specifically for the purposes of model validation. Some of these facilities have features such as movable thermal guard zones, the ability to cut off difficult-to-measure heat transfer paths, and detailing to minimize two- and three-dimensional conduction. Overall, it is best to design a test facility that can emulate the simplifying assumptions in the models, and then allow more realistic heat flows to be activated one at a time. In general, the more realistic the test building, the more difficult it is to establish causality and diagnose problems; the simpler and more controlled the test case, the easier it is to pinpoint sources of error or inaccuracy.

TASK 34/43

The objective of SHC Task 34/ECBCS Annex 43: Testing and Validation of Building Energy Simulation Tools was to undertake pre-normative research to continue to develop a comprehensive and integrated suite of building energy analysis tool tests that can provide software quality assurance. The parameter space within which building energy simulation software operates can be visualized by the matrix in Table 3 (Judkoff 2008). Each cell in the table represents an extremely large portion of the parameter space. IEA 34/43 sought to add tests to several of these cells as follows.

Comparative tests developed within IEA 34/43 included:

- BESTEST ground-coupled heat transfer with respect to floor slab construction (Project A)
- BESTEST multi-zone heat transfer, shading, and internal windows (Project B1)

SOFTWARE IMPROVEMENTS

This work has led to direct improvements in software tools used for evaluating the impacts of energy efficiency and solar energy technologies commonly applied in innovative low-energy buildings. During the field trials of the new test procedures, 106 results disagreements were identified, which led to 80 software fixes, including model and documentation improvements (other fixes are being worked on). Table 5 indicates by project the number of model errors that were identified and fixed so far. This indicates the utility of empirical validation (Projects C, D, and E), and analytical verification and comparative testing (Projects A and B) to identify disagreements that lead to corrections. Accuracy improvements in simulation models, demonstrated here and in previous IEA work, have increased confidence in model use by architects and engineers who rely on building energy simulation tool calculations to perform their work.
Project A also developed a new formal methodology to facilitate using and verifying numerical models to develop quasi-analytical solutions. This allows for greatly enhanced diagnostic capability when...
comparing results of other simplified and mid-level-detailed modeling methods that are typically used with whole-building energy simulation programs. Diagnostics are enhanced because the range of disagreement between quasi-analytical solutions for specific thermal mechanisms is typically much narrower than that between whole-building simulations. The methodology improvement also allows quasi-analytical solutions to be developed for more realistic (less constrained) cases than is possible for exact analytical solutions.

Table 5
Model Fixes Attributable to IEA SHC 34 / ECBCS Annex 43

<table>
<thead>
<tr>
<th>PROJECT</th>
<th>LEADER</th>
<th>DISAGreements</th>
<th>MODELS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Ground Coupled Slab-on-Grade, U.S./NREL</td>
<td>19</td>
<td>24</td>
<td>9</td>
</tr>
<tr>
<td>B2. Airflow, Japan</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>C. Shading/Daylighting/Load Interaction, Switzerland, U.S./Iowa</td>
<td>14</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>D. Mechanical Equipment and Controls, Germany</td>
<td>8</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>E2. Double-Skin Facade, Denmark</td>
<td>6</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>IEA SHC 34/ECBCS 43 TOTAL</td>
<td>80</td>
<td>106</td>
<td>24*</td>
</tr>
</tbody>
</table>

* Many models were tested in multiple projects as shown in Table 4

RESEARCH DISSEMINATION

The test procedures and final reports published by SHC Task 34/ECBCS Annex 43, summarized in Table 6, are posted at www.iea-shc.org/publications/task.aspx?Task=34. The participants have also published more than 20 conference papers and journal articles. Several more papers were developed for a special invited session on Annex 43 for the Building Simulation 2009 conference in Glasgow, UK, July 27-30, 2009, organized by the International Building Performance Simulation Association.

Table 6
Summary of Reports Developed in Annex 43

<table>
<thead>
<tr>
<th>PROJECT/TITLE</th>
<th>LEAD AUTHORS</th>
<th>LEAD COUNTRY</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. IEA BESTEST In-Depth Diagnostic Cases for Ground Coupled Heat Transfer Related to Slab-on-Grade Construction</td>
<td>J. Neymark, R. Judkoff</td>
<td>United States</td>
</tr>
<tr>
<td>B1. IEA BESTEST Multi-Zone Non-Airflow In-Depth Diagnostic Cases: MZ320 – MZ360</td>
<td>J. Neymark, R. Judkoff</td>
<td>United States</td>
</tr>
<tr>
<td>B2. Airflow Tests Including Multi-Zone Airflow</td>
<td>Y. Utsumi, T. Mitamura</td>
<td>Japan</td>
</tr>
<tr>
<td>C. Empirical Validations of Shading/Daylighting/Load Interactions in Building Energy Simulation Tools</td>
<td>P. Loutzenhiser, H. Manz, G. Maxwell</td>
<td>Switzerland, United States</td>
</tr>
<tr>
<td>D. Mechanical Equipment &amp; Control Strategies for a Chilled Water and a Hot Water System</td>
<td>C. Felsmann</td>
<td>Germany</td>
</tr>
<tr>
<td>E1. Double Skin Facades, a literature review</td>
<td>H. Poirazis</td>
<td>Sweden</td>
</tr>
<tr>
<td>E2. Empirical Validation of Building Simulation Software: Modeling of Double Facades</td>
<td>O. Kalyanova, P. Heiselberg</td>
<td>Denmark</td>
</tr>
</tbody>
</table>

Sources: Felsmann 2008; Loutzenhiser 2007; Neymark et al. 2008a, 2008b; Poirazis 2006

INDUSTRY USE OF BESTEST

National and international building energy standards organizations have used test cases developed in this Task and earlier ECBCS and SHC tasks to create standard methods of tests for building energy analysis tools. An example is ANSI/ASHRAE Standard 140, Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs which is used in the US for national building energy code compliance. The impact of such work is apparent from the following:

- Standard 140 is referenced by:
RECOMMENDATIONS AND CONCLUSIONS

As a result of SHC Task 34/ECBCS Annex 43, the content of the IEA Tool Evaluation Test Matrix has been augmented as shown in Table 3. This table also includes test procedures developed under SHC Task 12/ECBCS Annex 21, ECBCS Annex 42, and SHC Tasks 8 and 22. Although this represents a useful suite of basic tests, it does not address a broad enough cross-section of topics to represent a comprehensive evaluation of building energy analysis tools. Additional work should focus on filling missing or only partially covered areas of the matrix in Table 3.

Building energy simulation software must constantly be augmented to keep pace with new technology development. Thus, there is always a need for validation. A continuation of the validation work is recommended by the IEA 34/43 experts; additional tests could include, but are not limited to, those for models of the following:

- More HVAC system configurations
- More realistic building/ground-coupled heat transfer
- Model calibration methods for existing buildings (for predicting retrofit energy savings)
- Active solar thermal systems
- Thermo-chemical systems

REFERENCES


Kalyonava, O., Heiselberg, P. 2009. Double-Skin Façade Empirical Validation Tests at Aalborg University, Denmark [Draft: preliminary title], Aalborg University, Denmark.

**Title:** What Did They Do in IEA 34/43? Or How to Diagnose and Repair Bugs in 500,000 Lines of Code: Preprint

**Authors:** R. Judkoff and J. Neymark

**Performing Organization:**
National Renewable Energy Laboratory
1617 Cole Blvd.
Golden, CO 80401-3393

**Abstract:**
This paper presents an overview of the recently completed International Energy Agency Solar Heating and Cooling Task 34 and Energy Conservation in Buildings and Community Systems Annex 43 (IEA 34/43) on Testing and Validation of Building Energy Simulation Tools. The paper includes discussion of the technical and historical context, methodology, organization, accomplishments, and adoption by standards organizations and other regulatory entities.

**Subject Terms:**
iea; international energy agency; validation; whole-building energy simulation